

DNA Coding

<http://raic.kunsan.ac.kr>

로보틱스 및 인공지능제어 연구실

Robotics & Artificial Intelligent Control Laboratory

2002 / Designed by RAIC LAB. All rights reserved.

Content

1. 생물학적 DNA

1.1) DNA?

1.2) DNA의 역사

1.3) DNA의 구조

1.4) DNA의 역할

1.5) 용어 설명

2. DNA 코딩

1-1. DNA ?(1)

염색체는 당(deoxyribose)과 인산(phosphate)이 번갈아 가면서 연결된 긴 사슬 두 개와 이 사슬이 4종류의 염기(base)를 사이에 두고 나선처럼 꼬여 있는 이중나선구조(double helix strand)로 이루어져 있으며 이를 DNA(Deoxyribose Nucleic Acid)라고 합니다.

4종류의 염기는 아데닌(A: Adenine), 티민(T: Thymine), 구아닌(G: Guanine), 시토신(C: Cytosine)으로 아데닌(A)은 티민(T)과 구아닌(G)은 시토신(C)과 결합하도록 되어있으며, 인체의 경우 이들 염기가 짝을 이루는 조합수는 30억-31억개나 됩니다.

디지털 신호가 0과1로 구성되어 정보를 전달하듯 인간의 유전정보도 4종류의 염기 A, T, G, C가 특별한 배열 순서를 가짐으로써 인체구조와 생리활성을 맡는 단백질에 대한 유전 정보를 갖게 되는 것입니다.

1-1. DNA ?(2)

사람: 유전자는 DNA로 구성, 핵산 : DNA와 RNA로 구성되어 있다.

DNA가 단백질을 구성하는 데 필요한 화학정보를 가지고 있기는 하지만, DNA 자신이 직접 이와 같은 아미노산의 조립에 참가하는 것은 아니고, 이 화학 정보는 두 번째 종류의 핵산인 RNA로 전사된 후 이것이 아미노산을 지정해 단백질로 조립한다. 그러므로 우리가 가진 수천 개 유전자가 각기 그 기능을 나타내기 위해서는 먼저 RNA 분자가 형성되어야만 한다.

DNA는 자기 복제 능력이라는 독특한 기능을 가지고 있다. 이것은 DNA 자신이 자신과 똑같은 분자를 생산하는 것으로 자라는 생물에게 이 과정은 대단히 중요하다. 세포가 분열을 시작하기에 앞서 각 DNA 분자는 복제를 하게 되며, 세포 분열이 진행됨과 동시에 복제된 DNA는 정확하게 나뉘어 각기 새로 생겨난 두 개의 세포로 전달된다. 따라서 각 세대의 세포는 완전한 한 벌(set)의 유전자를 갖추게 된다.

1-2. DNA의 역사(1)

멘델의 유전 법칙 : 1865년 '식물의 교잡에 관한 실험'에서 유전법칙 발표
완두콩의 실험을 통한 확률적 법칙 발견.

1. 우열의 법칙: 생물의 특징을 나타내는 형질에는 우성과 열성이 있는데, 우성과 열성이 같이 있으면 우성만이 발현된다.
2. 분리의 법칙: 한 쌍의 대립유전자인 우성과 열성은 자손에 전해질 때 분리된다.
3. 독립의 법칙: 다른 형질을 나타내는 유전자들은 서로 독립적으로 행동한다. 예를 들면 완두콩의 모양을 결정하는 유전자와 색깔을 결정하는 유전자는 서로 독립적으로 작용한다

1-2. DNA의 역사(2)

1869(미셔, 독일)– DNA 발견

고름으로 부터 백혈구의 핵을 연구 ->뉴클레인 발견.

1880(코셀, 독일)– 세포 핵 물질의 주성분이 핵, 단백질이라는 것을 발견.

아데닌(A), 구아닌(G), 시토신(C), 티민(T), 우라실(U)
5가지의 염기 발견 및 명명.

DNA에서의 티민(T)는 RNA에서 우라실(U)로 표현된다.

1889(알트만, 독일)– 뉴클레인이 DNA와 단백질의 복합체라는 것을 밝힘.

염기, 인산, 당 만을 포함하는 물질을 뉴클레인과 구별
하기위하여 핵산(nucleic acid)라고 명명,
즉 DNA만을 따로 분리.

1944(에이버리, 영국)– DNA 분자가 자기 증식을 하여 그 특징을

다음세대에 물려준다는 것을 밝혀냄.

1-2. DNA의 역사(3)

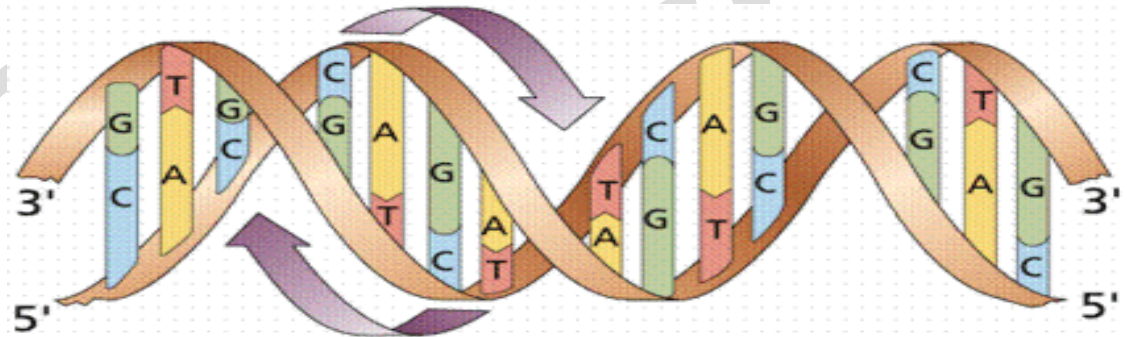
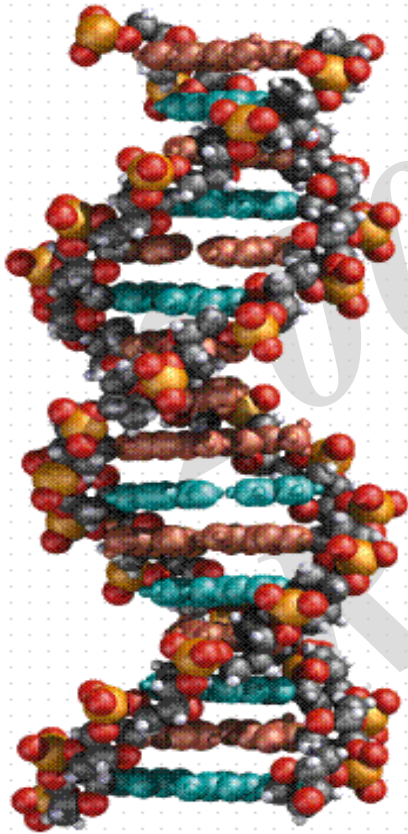
즉, 유전물질이 DNA라는 것을 밝힘.

미생물(박테리아, 바이러스)의 경우 유전 물질은 RNA이다.

1953(왓슨&크릭, 미국)– DNA의 구조가 이중나선임을 밝혀냄.

(분자 구조 모델 완성).

1-3. DNA의 구조(1)

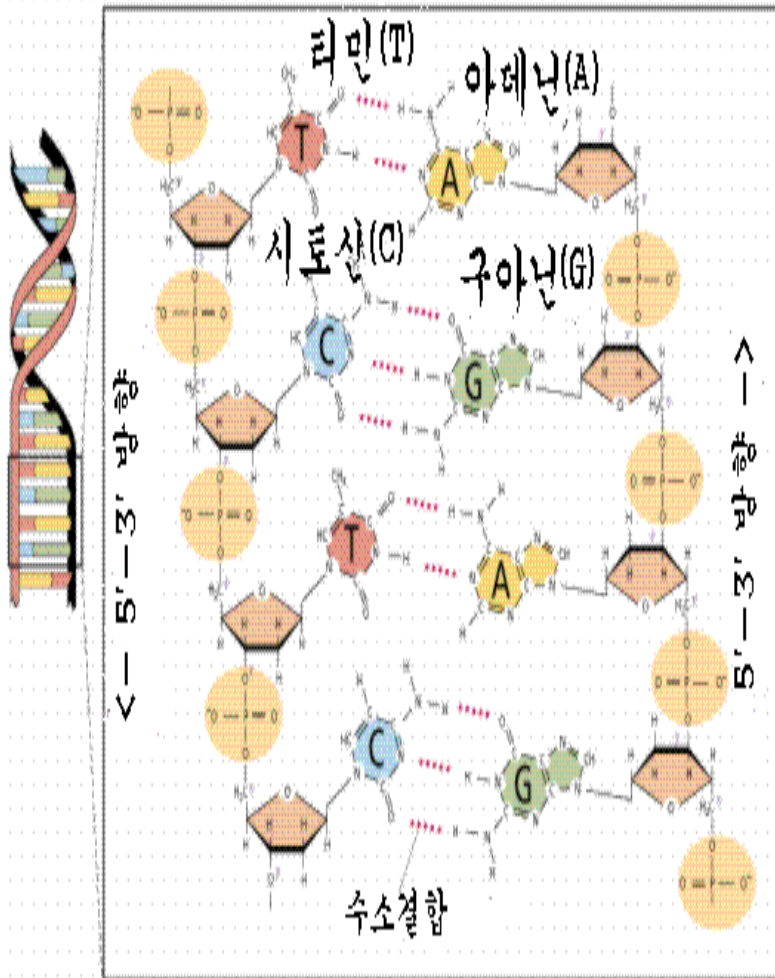


이중나선: 용수철을 잡아늘린 모양 두 개가 쌍을 이루면서 붙어있는 것을 말한다.

DNA가 하는 가장 중요한 일은 단백질을 만드는 것이다.

DNA가 RNA에게 정보를 전해 주면 그 RNA를 틀로 해서 단백질이 만들어진다는 것이다. 즉 DNA가 RNA를 낳고 RNA가 단백질을 낳는다는 것이다.

1-3. DNA의 구조(2)



DNA 이중나선을 만들 때에는 아데닌과 티민이 구아닌과 시토신이 수소결합을 하게 된다.

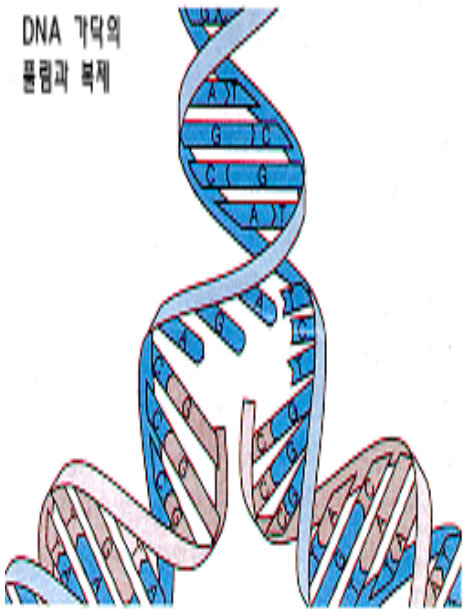
염기들의 수소결합을 통해서 두 개의 사슬이 연결되고 이중나선모양으로 꼬이게 되는 것이다.

DNA 이중나선은 히스톤단백질과 결합하여 뉴클레오솜으로 되고 뉴클레오솜으로 뭉쳐진 DNA 나선들이 직경 30nm 정도로 코일링된 것이 염색질이다. 이 염색질이 다시 코일링되면서 염색체로 된다.

핵속에서 DNA는 염색체의 상태로 있게 된다.

1-4. DNA의 역할 (1)

DNA 가닥의
분리과 복제

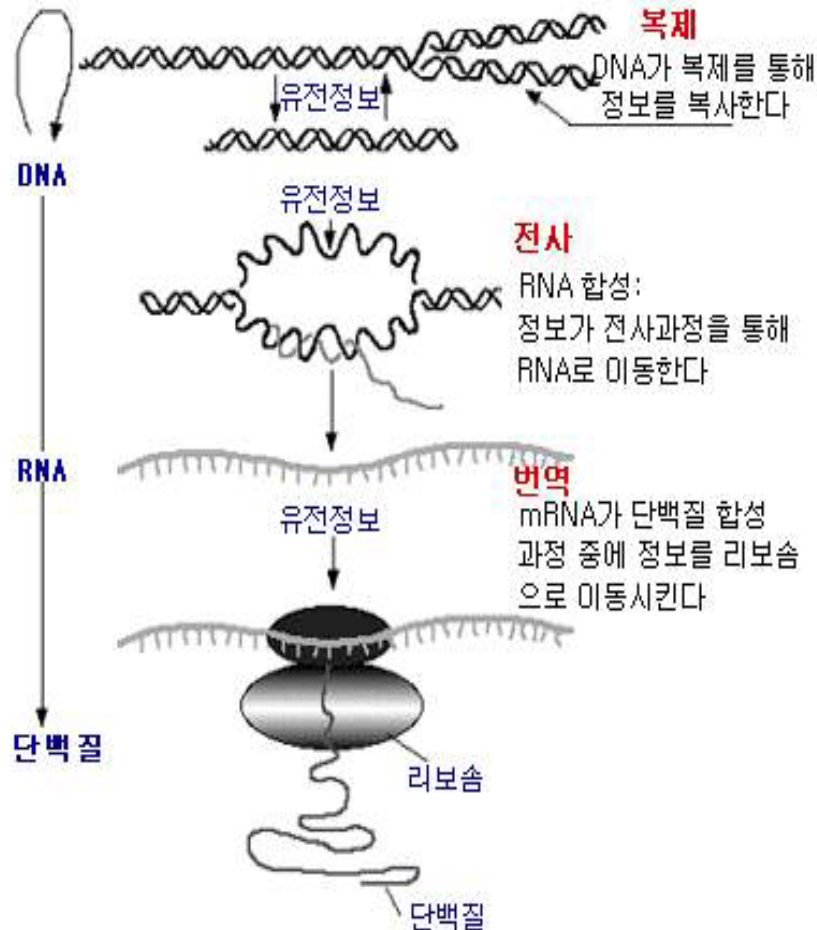


1. 자기 복제- 살아있는 이상 우리들의 몸 속 어딘가에서 세포가 분열하고있고 세포분열을 할 때에 원래 세포에 있던 지놈이 두 개로 되어야 한다. 예를들어, 혈액 중의 백혈구나 림프구는 몸을 순환하면서 몸을 지키는 세포로서 매일 새로 만들어져야 한다.

이들은 골수세포가 분열해서 만들어지는데 그것이 잘 안될 경우에 백혈병과 같은 혈액병이 생기기도 한다. 이럴 경우 외부로 부

세포를 집어 넣어주는 것이 골수이식이다.
간장 세포 또한 마찬가지이다.

1-4. DNA의 역할 (2)



2. 단백질을 만든다- 세포분열이 끝나면 각각의 장소에서 일을 해야 한다.
- 뇌세포가 뇌로서, 신경세포가 신경으로서 일하기 위해서는 거기에 필요한 단백질이 없으면 안 되는데 지놈이 바로 그 단백질을 만들고 있는 것이다.

1-4. DNA의 역할 (3)

3. 변한다- 진화론.

35억 년 전에 지구상에 최초로 태어난 생물의 지놈이 변하지 않았다면 지금도 최초의 생물이 그대로 살아 있을 것이다.

실제로 지금은 침팬지, 개, 고양이 등 어느 생명체도 자기의 지놈을 갖고 살고 있다. 지금 여러 가지 생명체가 갖고 있는 지놈을 거슬러 올라가면 최초의 지놈으로 부터 나온 것임에 틀림없다.

1-5. 용어 설명(1)

지놈, 유전자, DNA의 관계- 지놈은 세포에 있는 DNA 전부를 말하고,
지놈에 단백질을 만드는데 관계하고 있는 것이 유전자이다.
'유전자'라고 불리는 것은 주로 단백질을 합성하는 일을 하
고

있는 부분이지만, 그 이외에 '유전자'의 일을 조정하는
부분도 '유전자'라 불린다.

코돈- 단백질합성과정에서 어떤 아미노산을 만들지를 결정하는
세개의 뉴클레오타이드 로 된 염기배열.(64가지)

리보솜- 원형질내에서 볼 수 있는 구형의 입자로 단백질을 합성하는
곳이며, 60%의 RNA와 40%의 단백질로 구성되어 있다.

DNA 코딩

RNA(DNA) 코돈과 생성하는 아미노산

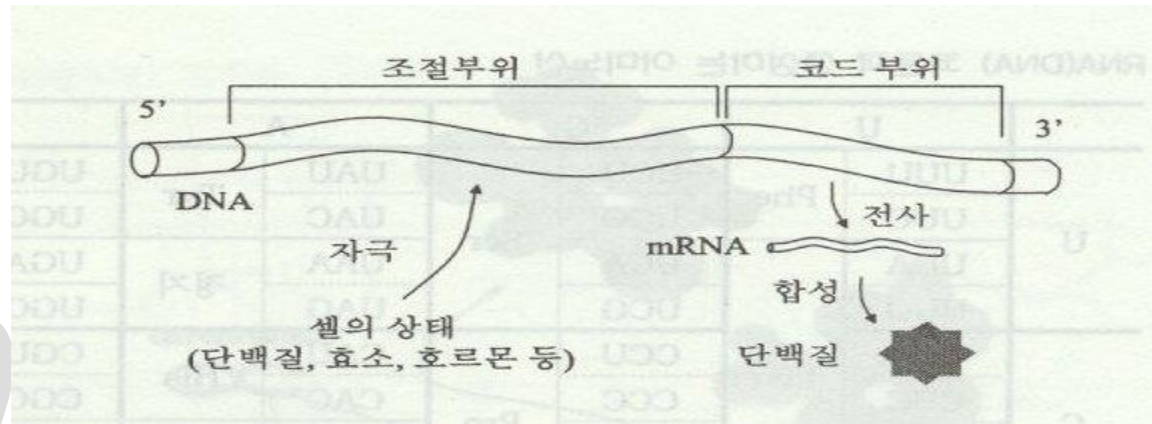
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	정지	UGA	정지	A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU		CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA	Met	ACA		AAA	Lys	AGA	Arg	A
	AUG		ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

아미노산 약어 알라닌-Ala, 아르기닌-Arg, 아스파라긴-Asn, 아스파르트산-Asp, 시스테인-Cys, 글루탐산-Glu, 글루타민-Gln, 글리신-Gly, 히스티딘-His, 이소루신-Ile, 루신-Leu, 리신-Lys, 메티오닌-Met, 페닐알라닌-Phe, 프롤린-Pro, 세린-Ser, 트레오닌-Thr, 트립토판-Trp, 티로신-Tyr, 발린-Val.

DNA 코딩

1. DNA는 RNA로 전사되어 리보솜에서 단백질로 번역된다. 즉 아미노산을 암호화 하는 DNA의 배열에 따라 아미노산의 합성 순서를 결정하여 여러 종류의 단백질을 만들어 낸다.
2. 그림은 RNA(DNA) 코돈과 생성하는 아미노산.
시작코돈 AUG에서부터 종료 코돈 UGA(UAA,UAG)까지 총 64개의 코돈이 존재하며, 이 64개의 코돈이 총 20가지의 아미노산을 생성한다.
코딩 부분에서는 20가지의 아미노산에 시작과 종료 코돈이 포함되어 20가지의 아미노산과 2가지의 코돈이 코딩된다.

DNA 코딩



1. 위의 그림은 일반적인 생물학적 유전자의 구조.

코드부위: 단백질을 직접 코드화.

조절부위: 코드 부위의 발현을 조절.

즉, 코드 부위는 조절부위의 명령에 의해 세포내의 조거나 조절부위를 자극할 때 단백질로 번역된다.

DNA	조절부위	코드부위
일반규칙	조건문(IF문)	결론부(Then절)

DNA 코딩

DNA코딩방법 순서.

1. 코딩을 위한 규칙의 표현방법 및 구성요소를 결정.
2. DNA코돈에 대응되는 규칙의 구성요소 테이블 작성.
3. 테이블을 이용해 번역하는 방법을 결정.
4. 진화 알고리즘을 이용하여 진화 시킴.

예) DNA코딩의 간단한 예(L-System 이용)

- L 시스템은 식물의 표현에 많이 사용된다.

간단한 L 시스템을 구성하기 위하여 문자는 줄기, 잎, 꽃을 나타내는 F, L, B의 3개 문자를 사용.

생성 규칙- 문맥 의존 및 Bracketed L-시스템을 사용

초기 문자열- 하나의 줄기 즉, F로 함.

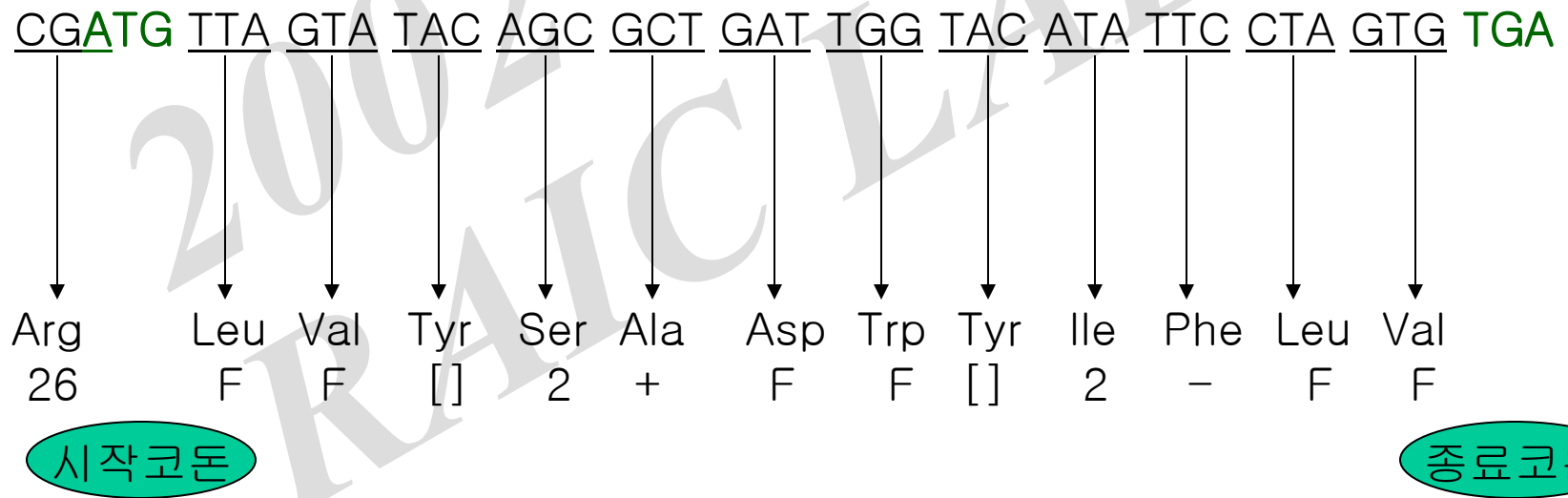
DNA 코딩

L-시스템의 코돈(아미노산) 번역표

Amino Acid	개수	각도(δ)	전건부	후건부	[]의 문자 수
Phe	2	65	F<B	—	5
Leu	6	30	F	F	0
Ile	3	50	F<F>B	B	2
Met	1	5	F<L	L	1
Val	4	20	F<F>F	F	0
Ser	6	35	B	B	2
Pro	4	40	F<F>L	L	1
Thr	4	28	F>F	[]	3
Ala	4	32	F>L	+	4
Tyr	2	55	F<F	[]	3
Stp	3	45			
His	2	85	F>L	+	4
Gln	2	60	F<L	+	4
Asn	2	80	B	B	2
Lys	2	10	F>B	—	5
Asp	2	70	F	F	0
Glu	2	15	F>F	[]	3
Cys	2	75	L	L	1
Trp	1	90	F<F	F	0
Arg	6	26	L	L	1
Gly	4	24	F>B	—	5

DNA 코딩

L- 시스템의 염색체 번역을 이용한 예)



DNA 코딩

1. 염색체의 첫 코돈: 식물의 가지가 뻗는 각도(δ)
2. 시작코돈: ATG
3. 첫번째 코돈: 전건부
4. 두번째 코돈~: 후건부
5. 종료코돈: TGA

실질적인 번역.

1. CGA(Arg) 각도=26(CGA=Arg=26)
2. ATG: 시작 코돈, TTA(Leu): 전건부 F
3. GTA(Val): 후건부 F~,
4. TGA: 종료 코돈
5. 규칙생성: $F \rightarrow F[+F]F[-F]F$

DNA 코딩

DNA코딩의 특징: 발생 모델의 진화 표현에 효과적.

염색체는 기본적으로 4가지 염기의 배열로 이루어져 있고,
아미노산을 번역하는 것과 마찬가지로 코돈 단위로 번역.

1. 중복 번역: 시작점이 일정하지 않기 때문.

중복 번역으로 인해 번역 되지 않는 부분도 있다.

2. 여러 개의 코돈이 하나의 아미노산을 만드는 특성: 염색체의 중복을
효율적으로 이용할 수 있다.

3. 염색체의 길이 가변적: 두 부모 염색체의 교차점이 별도로 지정되기
때문에 길이 가변적.

DNA 코딩

생물학적 DNA와 DNA코딩의 비교.

	생물학적 DNA	DNA코딩
코돈	아미노산을 암호화 하는 최소 단위	규칙의 최소 의미 단위
시작 코돈	번역의 시작점	번역의 시작점
종료 코돈	번역의 종료점	번역의 종료점
번역 결과물	단백질, 효소	규칙

Bit String코딩에 비해 DNA코딩의 특징.

- 염색체의 길이가 가변적.
- 교차점이 강제적으로 주어지지 않는다.
- 코딩의 여분과 중복이 있다.
- 지식의 융통성 있는 표현(규칙)이 가능하다.